

Competing risks

Manuel Rain

14-11-2020

Contents

Resumen	1
Ejemplo	2
Comparación no paramétrica de CIF	2
Regresión de peligros por causas específicas	3
Modelo de peligros de subdistribución (SH)	6
Predicción del modelo	8
Diagnóstico del modelo	10
Referencia	14

Resumen

El análisis de supervivencia en presencia de riesgos en competencia impone desafíos adicionales para los investigadores clínicos en el sentido de que la función de riesgo (la tasa) no tiene un vínculo uno a uno con la función de incidencia acumulada (CIF, el riesgo). CIF es de particular interés y se puede estimar de forma no paramétrica con la función `useCUMINC()`. Esta función también permite la comparación de grupos y la visualización del CIF estimado. El efecto de las covariables sobre el peligro de causa específica se puede explorar utilizando el modelo convencional de riesgo proporcional de Cox al tratar los eventos en competencia como censura. Sin embargo, el efecto sobre el peligro no puede vincularse directamente con el efecto sobre los CIF porque no existe una correspondencia uno a uno entre el peligro y la incidencia acumulada. El modelo Fine-Gray modela directamente el efecto de la covariable en CIF e informa el índice de riesgo de subdistribución (SHR). Sin embargo, SHR solo proporciona información sobre el orden de las curvas CIF en diferentes niveles de covariables, no tiene una interpretación práctica como HR en ausencia de riesgos competitivos. El modelo `grsfr` se puede ajustar con la función `crr()` enviada con el paquete `cmprsk`. Se permiten covariables que varían en el tiempo en la función `crr()`, que se especifica mediante los argumentos `cov2` y `tf`. Se permiten predicciones y visualización de CIF para sujetos con valores de covariables dados para el objeto `crr`. Alternativamente, los modelos de riesgo en competencia pueden ajustarse al paquete `riskRegression` mediante el empleo de diferentes funciones de enlace entre covariables y resultados. El supuesto de proporcionalidad se puede verificar probando la significancia estadística de los términos de interacción que involucran tiempo de falla. Los residuos de Schoenfeld proporcionan otra forma de verificar el supuesto del modelo.

Ejemplo

El ejemplo trabajado está contenido en el paquete `riskRegression` . Un total de 205 pacientes con melanoma se habían sometido a una operación quirúrgica y se les dio seguimiento hasta finales de 1977. El conjunto de datos se puede cargar y visualizar de la siguiente manera.

```
library(riskRegression)
library(ggfortify)
library(survminer)
library(plotly)
library(gridExtra)

data(Melanoma)
set.seed(234)
Melanoma$id=sample(1:205)

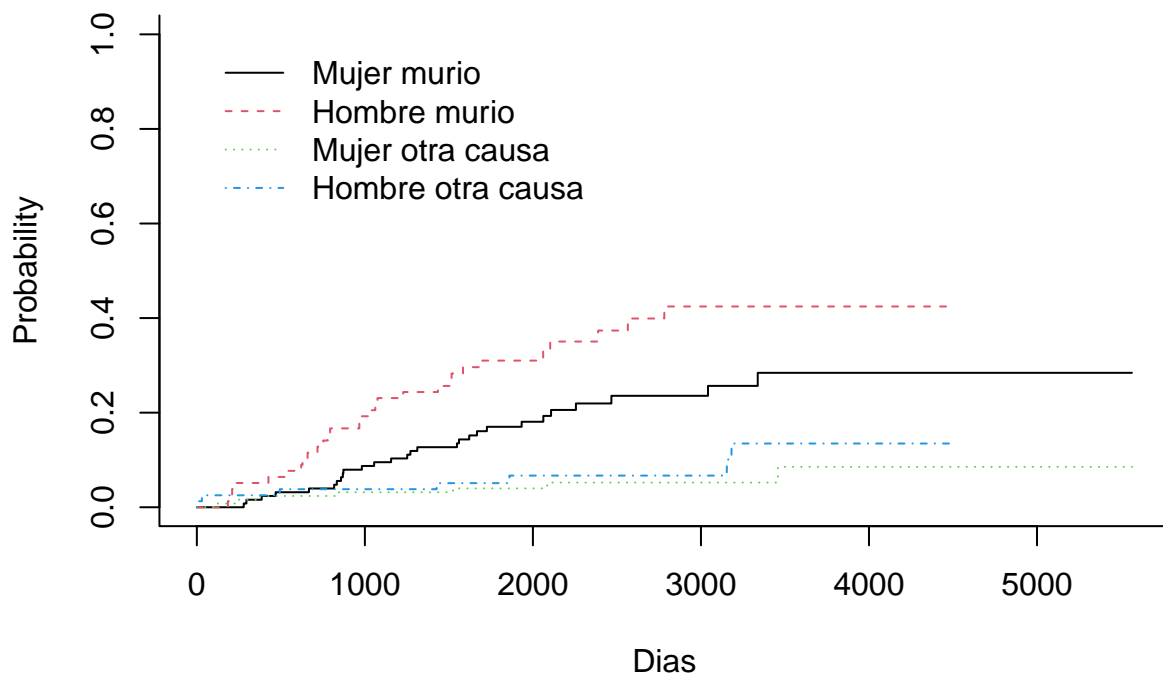
Melanoma$event=factor(Melanoma$event, levels = c("censored",
                                                  "death.malignant.melanoma",
                                                  "death.other.causes"),
                      labels = c("censurado", "murio",
                                "otra causa"))

Melanoma$sex=factor(Melanoma$sex, levels = c("Female", "Male"),
                   labels = c("Mujer", "Hombre"))
```

Comparación no paramétrica de CIF

Se pueden emplear CIF para diferentes causas de falla para la descripción estadística de datos de supervivencia con riesgos competitivos. Por lo tanto, los CIF para diferentes causas de falla brindan información adicional sobre los datos de supervivencia disponibles. La función `cuminc()` incluida con el paquete `cmprsk` puede estimar los CIF para diferentes causas de falla y permite comparaciones entre grupos.

```
library(cmprsk )
cif<-cuminc(ftime = Melanoma$time, fstatus = Melanoma$event,
           group=Melanoma$sex, cencode = "censurado")
plot(cif,col=1:5,xlab="Dias")
```



La Figura muestra que los pacientes varones tienen un mayor riesgo de muerte por melanoma y otras causas que las mujeres. La diferencia parece mayor para el fracaso por melanoma versus el fracaso por otras causas.

```
cif$Test
```

```
##           stat      pv df
## murio      5.8140209 0.0158989 1
## otra causa 0.8543656 0.3553203 1
```

La primera columna del resultado muestra el estadístico χ^2 para la prueba entre grupos, y la segunda columna muestra los valores P respectivos. Los pacientes masculinos tienen más probabilidades de morir de melanoma que las mujeres ($P = 0,016$), pero no hay una diferencia significativa en el riesgo de mortalidad por otras causas para los pacientes masculinos y femeninos ($P = 0,36$).

Regresión de peligros por causas específicas

El modelo de regresión de peligros específicos de la causa se puede ajustar con la regresión de Cox al tratar las fallas de la causa de interés como eventos y las fallas de otras causas como observación censurada. El efecto de las covariables sobre el riesgo de causa específica se puede estimar con la regresión de riesgo proporcional de COX.

```
csh<-coxph(Surv(time,event=="murio")~sex+age+invasion,data=Melanoma)
summary(csh)
```

```
## Call:
## coxph(formula = Surv(time, event == "murio") ~ sex + age + invasion,
##       data = Melanoma)
##
```

```
## n= 205, number of events= 57
##
##               coef exp(coef) se(coef)      z Pr(>|z|)
## sexHombre      0.663383  1.941349 0.266320 2.491 0.012741 *
## age             0.009823  1.009871 0.008339 1.178 0.238840
## invasionlevel.1 1.037168  2.821217 0.328241 3.160 0.001579 **
## invasionlevel.2 1.403225  4.068300 0.380744 3.685 0.000228 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##               exp(coef) exp(-coef) lower .95 upper .95
## sexHombre          1.941      0.5151   1.1519   3.272
## age                1.010      0.9902   0.9935   1.027
## invasionlevel.1     2.821      0.3545   1.4826   5.368
## invasionlevel.2     4.068      0.2458   1.9290   8.580
##
## Concordance= 0.7 (se = 0.035 )
## Likelihood ratio test= 26.7 on 4 df,  p=2e-05
## Wald test              = 24.39 on 4 df,  p=7e-05
## Score (logrank) test = 26.85 on 4 df,  p=2e-05
```

Alternativamente, la tarea se puede realizar usando la función `CSC()` contenida en el paquete `riskRegression`.

```
library("riskRegression")
library("prodlim")
```

```
CSH<-CSC(Hist(time,event)~sex+age+invasion,data=Melanoma)
CSH
```

```
## CSC(formula = Hist(time, event) ~ sex + age + invasion, data = Melanoma)
##
## Uncensored response of a competing.risks model
##
## No.Observations: 205
##
## Pattern:
##
## Cause      event
## censurado   134
## murio        57
## otra causa   14
## unknown      0
##
##
## -----> Cause: censurado
##
## Call:
## survival::coxph(formula = survival::Surv(time, status) ~ sex +
##   age + invasion, x = TRUE, y = TRUE)
##
## n= 205, number of events= 134
##
##               coef exp(coef) se(coef)      z Pr(>|z|)
## sexHombre      -0.074476  0.928230 0.189559 -0.393 0.69440
```

```

## age          0.019687  1.019882  0.006526  3.017  0.00255 **
## invasionlevel.1 -0.039552  0.961220  0.202450 -0.195  0.84511
## invasionlevel.2 -0.210257  0.810376  0.285327 -0.737  0.46118
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##               exp(coef) exp(-coef) lower .95 upper .95
## sexHombre      0.9282      1.0773      0.6402      1.346
## age            1.0199      0.9805      1.0069      1.033
## invasionlevel.1  0.9612      1.0403      0.6464      1.429
## invasionlevel.2  0.8104      1.2340      0.4633      1.418
##
## Concordance= 0.579 (se = 0.029 )
## Likelihood ratio test= 11.06 on 4 df,  p=0.03
## Wald test              = 10.5 on 4 df,  p=0.03
## Score (logrank) test = 10.59 on 4 df,  p=0.03
##
##
## -----> Cause:  murio
##
## Call:
## survival::coxph(formula = survival::Surv(time, status) ~ sex +
##   age + invasion, x = TRUE, y = TRUE)
##
##   n= 205, number of events= 57
##
##               coef exp(coef) se(coef)      z Pr(>|z|)
## sexHombre      0.663383  1.941349  0.266320  2.491 0.012741 *
## age            0.009823  1.009871  0.008339  1.178 0.238840
## invasionlevel.1 1.037168  2.821217  0.328241  3.160 0.001579 **
## invasionlevel.2 1.403225  4.068300  0.380744  3.685 0.000228 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##               exp(coef) exp(-coef) lower .95 upper .95
## sexHombre      1.941      0.5151      1.1519      3.272
## age            1.010      0.9902      0.9935      1.027
## invasionlevel.1  2.821      0.3545      1.4826      5.368
## invasionlevel.2  4.068      0.2458      1.9290      8.580
##
## Concordance= 0.7 (se = 0.035 )
## Likelihood ratio test= 26.7 on 4 df,  p=2e-05
## Wald test              = 24.39 on 4 df,  p=7e-05
## Score (logrank) test = 26.85 on 4 df,  p=2e-05
##
##
## -----> Cause:  otra causa
##
## Call:
## survival::coxph(formula = survival::Surv(time, status) ~ sex +
##   age + invasion, x = TRUE, y = TRUE)
##

```

```
## n= 205, number of events= 14
##
##              coef exp(coef) se(coef)      z Pr(>|z|)
## sexHombre      0.29893   1.34842  0.54620  0.547 0.584180
## age             0.09271   1.09715  0.02592  3.576 0.000349 ***
## invasionlevel.1 -0.88527   0.41260  0.64069 -1.382 0.167051
## invasionlevel.2 -1.34958   0.25935  1.12351 -1.201 0.229666
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##              exp(coef) exp(-coef) lower .95 upper .95
## sexHombre      1.3484      0.7416   0.46227   3.933
## age             1.0971      0.9115   1.04279   1.154
## invasionlevel.1  0.4126      2.4236   0.11754   1.448
## invasionlevel.2  0.2593      3.8558   0.02868   2.345
##
## Concordance= 0.839 (se = 0.042 )
## Likelihood ratio test= 18.96 on 4 df,  p=8e-04
## Wald test              = 14.21 on 4 df,  p=0.007
## Score (logrank) test = 15.53 on 4 df,  p=0.004
```

La salida resumida es bastante similar a la producida por la función `coxph()` excepto que la función `CSC()` produce automáticamente modelos de peligro de causa específica para ambos tipos de eventos (causa murio y murio otra causa). Con el modelo de regresión ajustado, se puede predecir el riesgo individual con covariables dadas.

Por ejemplo, quiero predecir el riesgo de un paciente masculino de 40 años con nivel de invasión 2.

```
library("pec")

pec::predictEventProb(CSH,cause="murio",
  newdata=data.frame(age=40,invasion=factor("level.2",levels=levels(Melanoma$invasion)),
    sex=factor("Hombre",levels=levels(Melanoma$sex))),
  time=c(1000,2000,3000))

##           [,1]      [,2]      [,3]
## [1,] 0.2938223 0.4961289 0.5977142
```

Modelo de peligros de subdistribución (SH)

El modelo SHs también se conoce como modelo Fine-Grey. Es un modelo de regresión proporcional de Cox, pero la incidencia acumulada está asociada con las HS. La motivación del modelo de Fine-Gray es que el efecto de una covariable en la función de riesgo de causa específica puede ser bastante diferente al de CIF. En otras palabras, una covariable puede tener una gran influencia en la función de riesgo de causa específica, pero no tiene ningún efecto en el CIF. La diferencia entre el peligro por causa específica y la subdistribución es que los eventos de riesgo en competencia se tratan de manera diferente. El primero considera los eventos de riesgo en competencia como una censura no informativa, mientras que el segundo toma en cuenta la naturaleza de censura informativa de los eventos de riesgo en competencia.

```
SH <- FGR(Hist(time,event=="murio")~sex+age+invasion,data=Melanoma)
SH

##
## Right-censored response of a survival model
##
## No.Observations: 205
```

```
##
## Pattern:
##           Freq
## event      57
## right.censored 148
##
##
## Fine-Gray model: analysis of cause 1
##
## Competing Risks Regression
##
## Call:
## FGR(formula = Hist(time, event == "murio") ~ sex + age + invasion,
##      data = Melanoma, cause = "1")
##
##           coef exp(coef) se(coef)      z p-value
## sexHombre    0.62288     1.86  0.27035  2.30 0.02100
## age           0.00326     1.00  0.00906  0.36 0.72000
## invasionlevel.1 1.00836     2.74  0.33825  2.98 0.00290
## invasionlevel.2 1.38232     3.98  0.40078  3.45 0.00056
##
##           exp(coef) exp(-coef)  2.5% 97.5%
## sexHombre         1.86      0.536 1.097  3.17
## age                1.00      0.997 0.986  1.02
## invasionlevel.1     2.74      0.365 1.413  5.32
## invasionlevel.2     3.98      0.251 1.816  8.74
##
## Num. cases = 205
## Pseudo Log-likelihood = -283
## Pseudo likelihood ratio test = 23.1 on 4 df,
##
## Convergence: TRUE
```

Como puede ver, el coeficiente estimado para la causa murio se desvía un poco del obtenido del modelo de riesgo de causa específica (HR: 1,87 frente a 1,94), lo que refleja diferentes supuestos para los riesgos en competencia. Los valores numéricos derivados del modelo Fine-Gray no tienen una interpretación simple, pero reflejan el orden de las curvas de incidencia acumulada. El riesgo de causa específica es la tasa de falla de causa murio por unidad de tiempo para pacientes que aún están vivos. Sin embargo, la causa murio SH es la tasa de falla de la causa 1 por unidad de tiempo para los pacientes que están vivos o ya han fallado por la causa 2. En otras palabras, los pacientes que fallan por otras causas todavía están en el grupo de riesgo.

El modelo Fine-Grey se puede ajustar con la función `crr()` en el paquete `cmprsk`. Los argumentos de la función son diferentes de los de la función `FGR()`. Aunque no se admite el uso de la fórmula del modelo, la función `model.matrix` se puede utilizar para generar matrices adecuadas de covariables a partir de factores.

```
cov<-model.matrix(~sex+age+invasion,data=Melanoma)[-1]
crr.model<-crr(Melanoma$time,Melanoma$status,cov1=cov)
crr.model
```

```
## convergence: TRUE
## coefficients:
##      sexHombre      age invasionlevel.1 invasionlevel.2
##      0.627600      0.005653      1.049000      1.378000
## standard errors:
## [1] 0.271700 0.009131 0.340400 0.401400
## two-sided p-values:
```

```
##      sexHombre      age invasionlevel.1 invasionlevel.2
##      0.0210      0.5400      0.0021      0.0006
```

Predicción del modelo

El modelo ajustado de Fine-Gray se puede utilizar para predecir nuevas observaciones con combinaciones dadas de covariables. En el siguiente ejemplo, se proporciona un nuevo conjunto de datos que contiene tres pacientes. Se definen para ellos covariables de edad, sexo y niveles de invasión

```
newdata<-data.frame(sex=factor(c("Hombre","Hombre","Mujer"),
                             levels=levels(Melanoma$sex)),age=c(52,32,59),
                    invasion=factor(c("level.2","level.1","level.2"),
                                    levels=levels(Melanoma$invasion)))
```

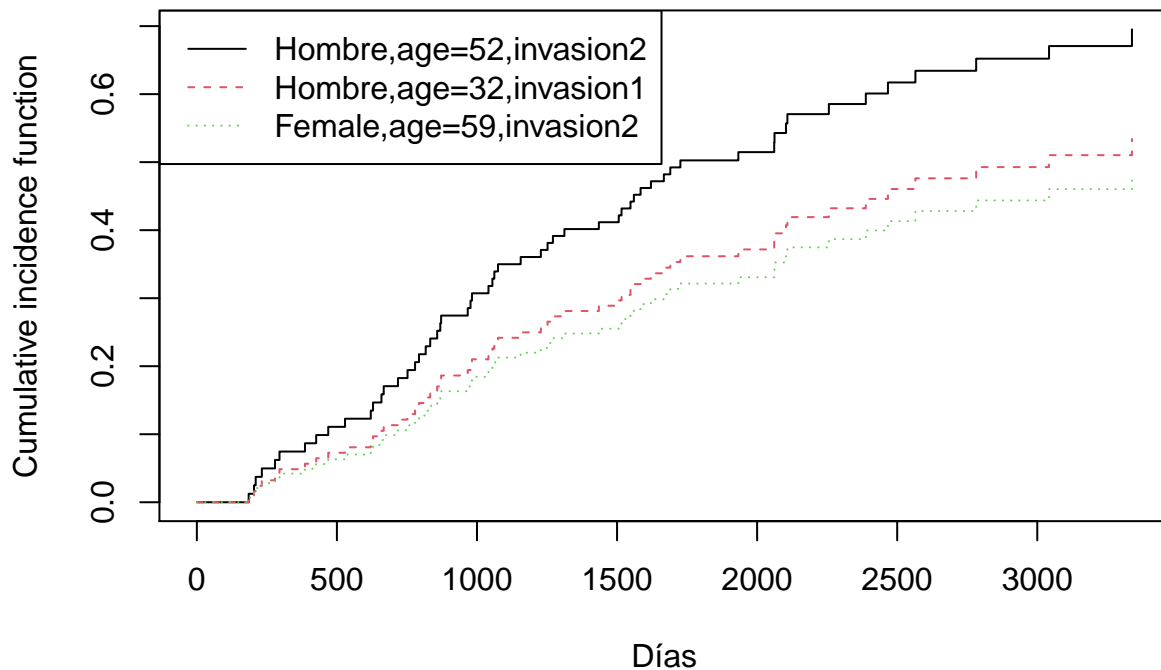
```
newdata
```

```
##      sex age invasion
## 1 Hombre  52 level.2
## 2 Hombre  32 level.1
## 3  Mujer  59 level.2
```

```
dummy.new<-model.matrix(~sex+age+invasion,data=newdata)[,-1]
dummy.new
```

```
##      sexHombre age invasionlevel.1 invasionlevel.2
## 1      1    52      0      1
## 2      1    32      1      0
## 3      0    59      0      1
```

```
pred<-predict(crr.model,dummy.new)
plot(pred,lty=1:3,col=1:3,xlab="Días",ylab="Cumulative incidence function")
legend("topleft",c("Hombre,age=52,invasion2",
                  "Hombre,age=32,invasion1","Female,age=59,invasion2"),
      lty=1:3,col=1:3)
```

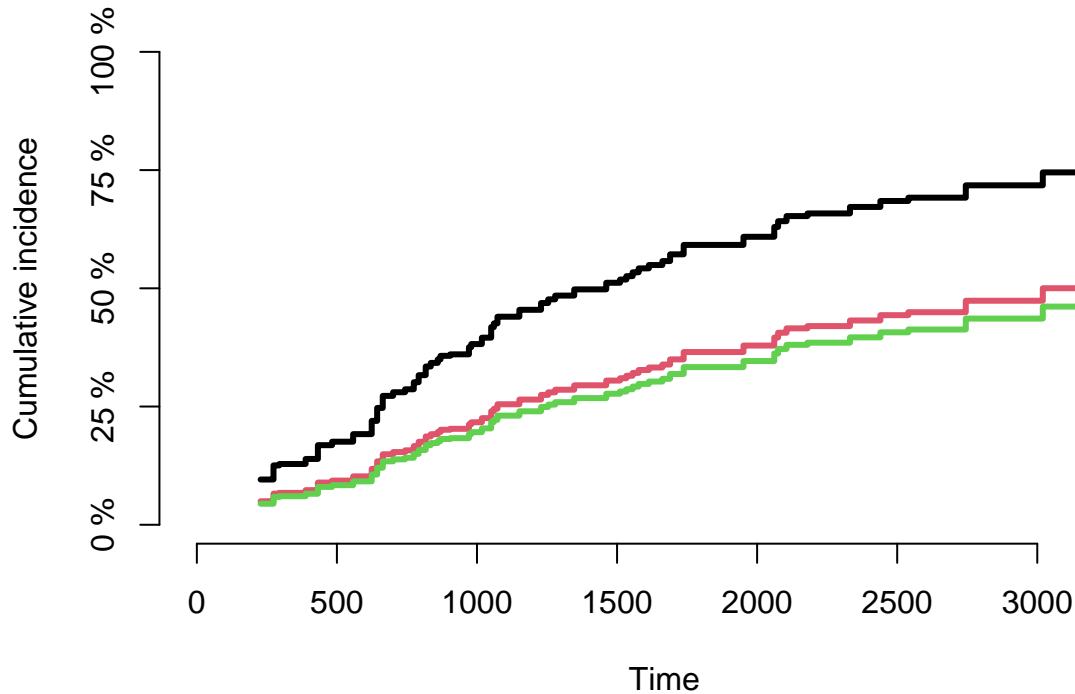



La salida gráfica anterior proporciona CIF para pacientes con características especificadas en los nuevos datos . El argumento de enlace controla la función de enlace que se utilizará: “prop” para el modelo de regresión de Fine-Gray, “relativo” para el modelo de regresión de riesgo absoluto y “logístico” para el modelo de regresión de riesgo logístico.

```
reg<-riskRegression(Hist(time, status) ~ sex + age +invasion,
                    data = Melanoma, cause = 1,link="prop")
reg
```

```
## Competing risks regression model
##
## IPCW weights: marginal Kaplan-Meier for the censoring distribution.
## Link: 'cloglog' yielding sub-hazard ratios (Fine & Gray 1999)
## No covariates with time-varying coefficient specified.
##
## Time constant regression coefficients:
##   Variable  Levels Coef Lower Upper  Pvalue
##      sex    Hombre 2.23 1.255  3.95 0.00622
##      age              1.00 0.983  1.02 0.90659
##  invasion level.1 2.64 1.312  5.33 0.00656
##              level.2 5.10 2.331 11.14 < 1e-04
##
##
## Note: The coefficients (Coef) are sub-hazard ratios (Fine & Gray 1999)
```

```
plot(reg,newdata=newdata)
```



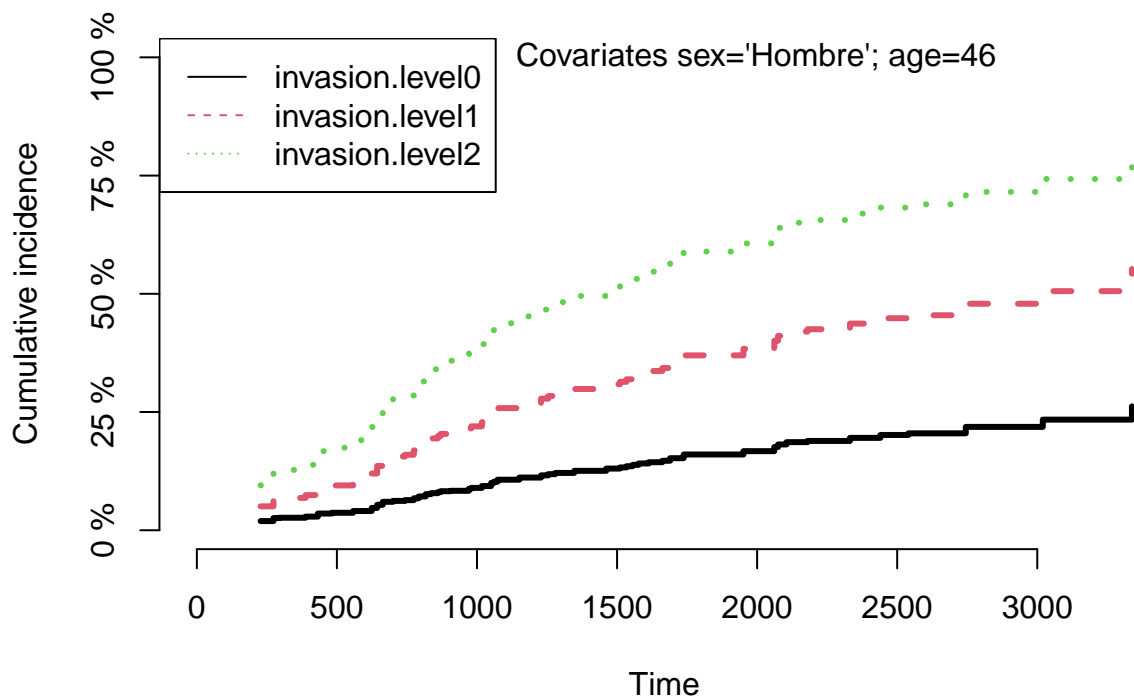
Diagnóstico del modelo

Un supuesto importante del modelo de regresión de Cox es la proporcionalidad, que supone que la subdistribución con covariables z es un cambio constante en la escala logarítmica complementaria de una función de subdistribución de referencia. Las curvas no se cruzarán entre sí. La verificación del modelo puede realizarse inicialmente mediante un examen gráfico de los CIF.

```
checkdata<-data.frame(sex=factor(c("Hombre","Hombre","Hombre"),
                                levels=levels(Melanoma$sex)),
                      age=c(46,46,46),
                      invasion=factor(c("level.0","level.1","level.2"),
                                       levels=levels(Melanoma$invasion)))
checkdata
```

```
##      sex age invasion
## 1 Hombre  46 level.0
## 2 Hombre  46 level.1
## 3 Hombre  46 level.2
```

```
plot(reg,newdata=checkdata,lty=1:3,col=1:3)
text(2000,1,"Covariates sex='Hombre'; age=46")
legend("topleft",c("invasion.level0","invasion.level1","invasion.level2"),
      lty=1:3,col=1:3)
```



La Figura muestra los CIF en diferentes niveles de invasión, estableciendo la edad en 46 años y el sexo en el hombre. No hay evidencia de violación del supuesto de proporcionalidad para la invasión variable. Otro método para verificar el supuesto proporcional es incluir una covariable dependiente del tiempo en el modelo de regresión.

```
crr.time<-crr(Melanoma$time,Melanoma$status,cov1=cov,
               cov2=cov[,1],tf=function(t) t)
summary(crr.time)
```

```
## Competing Risks Regression
```

```
##
```

```
## Call:
```

```
## crr(ftime = Melanoma$time, fstatus = Melanoma$status, cov1 = cov,
```

```
##     cov2 = cov[, 1], tf = function(t) t)
```

```
##
```

```
##               coef exp(coef) se(coef)      z p-value
```

```
## sexHombre      1.144721      3.14 0.514436  2.225 0.02600
```

```
## age             0.005667      1.01 0.009013  0.629 0.53000
```

```
## invasionlevel.1 1.049050      2.85 0.338726  3.097 0.00200
```

```
## invasionlevel.2 1.375556      3.96 0.397515  3.460 0.00054
```

```
## cov[, 1]*tf1    -0.000413      1.00 0.000349 -1.182 0.24000
```

```
##
```

```
##               exp(coef) exp(-coef)  2.5% 97.5%
```

```
## sexHombre          3.14      0.318 1.146  8.61
```

```
## age                1.01      0.994 0.988  1.02
```

```
## invasionlevel.1     2.85      0.350 1.470  5.55
```

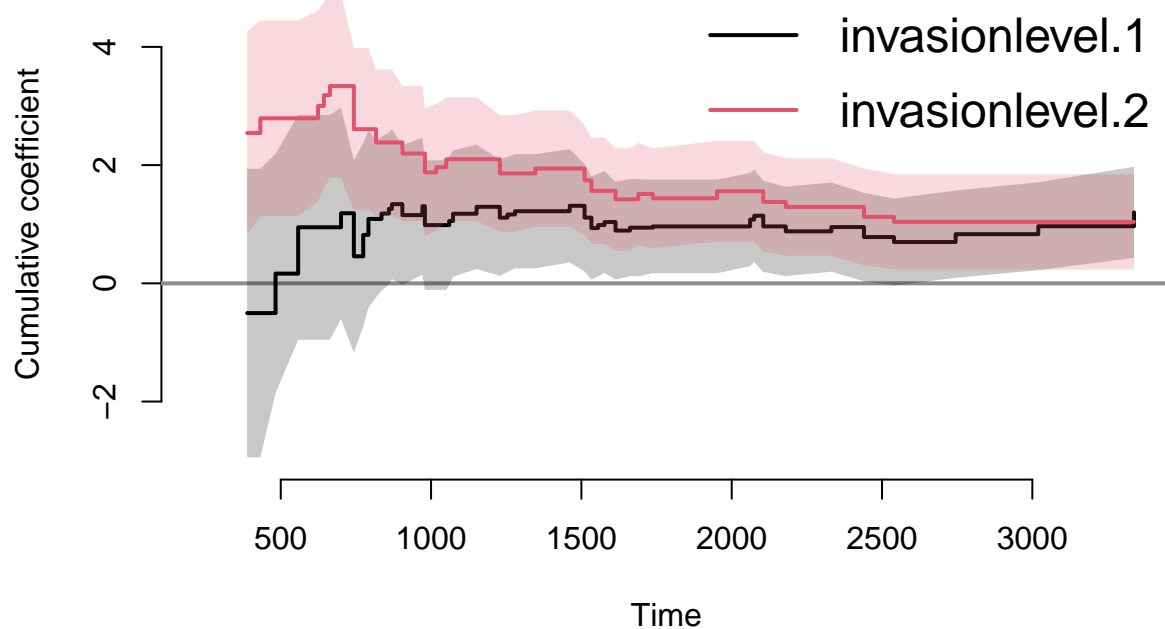
```
## invasionlevel.2     3.96      0.253 1.816  8.63
```

```
## cov[, 1]*tf1      1.00      1.000 0.999 1.00
##
## Num. cases = 205
## Pseudo Log-likelihood = -273
## Pseudo likelihood ratio test = 25.9 on 5 df,
```

El argumento cov2 toma una matriz de covariables que se multiplicará por el tiempo. Las funciones del tiempo se especifican en el argumento tf. La función toma un vector de tiempos como argumento y devuelve una matriz. La j-ésima columna de la matriz de tiempo se multiplicará por la j-ésima columna de cov2. Por ejemplo, un modelo de la forma se puede especificar en la función crr () mediante (cov1 = x 1, cov2 = cbind (x 1, x 1), tf = function (t) cbind (t, t ^ 2)). En el resultado resumido del modelo Fine-Gray con covariable variable en el tiempo, el último término no muestra significación estadística (P = 0,24), lo que indica que el efecto del sexo es constante en el tiempo. El riesgo de regresión proporciona una resolución simple para modelar la covariable variable en el tiempo.

```
reg.time<-riskRegression(Hist(time, status) ~ sex + age +
                        strata(invasion),
                        data = Melanoma, cause = 1,link="prop")

plotEffects(reg.time,formula=~invasion)
```

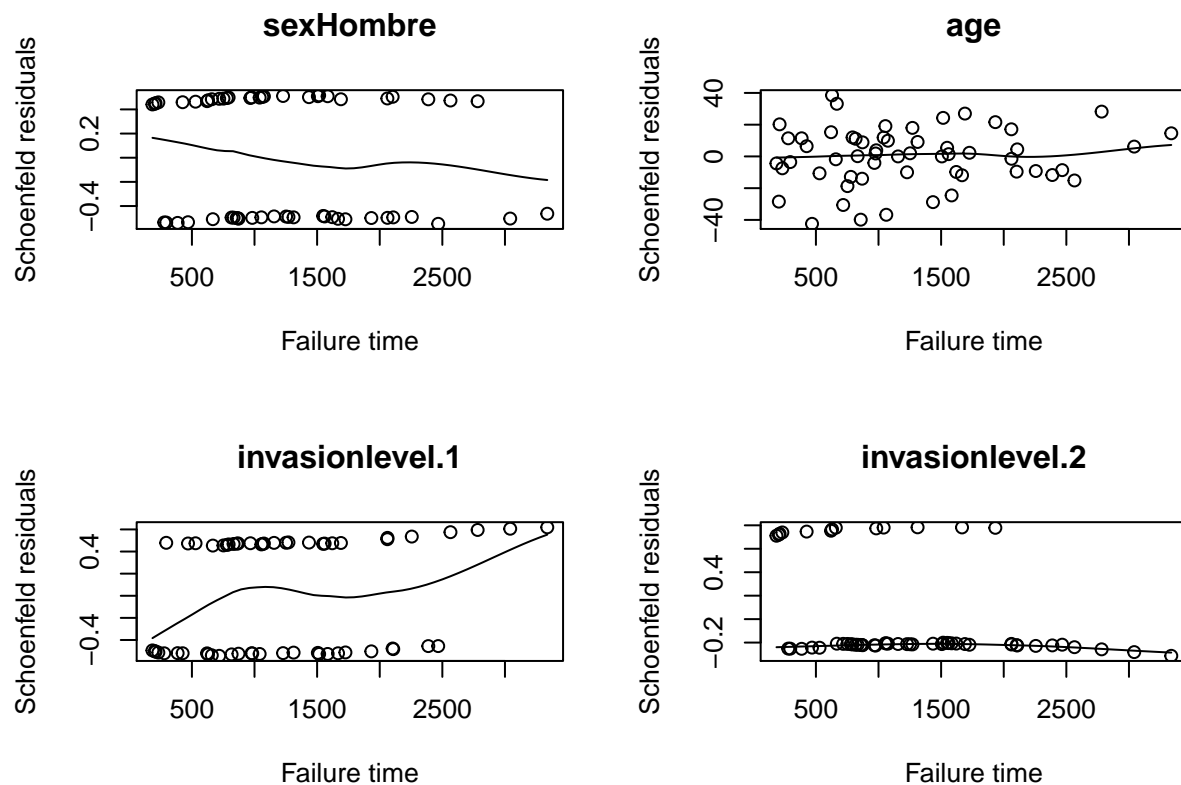


La Grafica muestra los efectos dependientes del tiempo en el modelo de regresión de Fine-Gray para la mortalidad. La curva y el intervalo de confianza del 95% correspondiente se trazan con un método no paramétrico. Parece que los coeficientes para el nivel 2 vs. 1 son más grandes durante el período de tiempo de 0 a 1000 que en otros momentos, lo que indica algunas interacciones temporales leves (12). Sin embargo, la prueba estadística formal no está permitida en este entorno.

Trazo los residuos de Schoenfeld contra el tiempo de falla para cada covariable. Si el supuesto proporcional es

cierto, el residuo debe tener una media constante a lo largo del tiempo. Se agrega un diagrama de dispersión más suave para cada covariable para verificar el supuesto (Figura 5). Parece que el nivel de invasión 1 tiene residuos no constantes a lo largo del tiempo, lo que indica una posible violación del supuesto proporcional. Para comprobar formalmente la suposición del nivel de invasión 1, podemos agregar un término de interacción con el tiempo.

```
par(mfrow=c(2,2))
for(j in 1:ncol(crr.model$res)) {
  scatter.smooth(crr.model$uftime, crr.model$res[,j],
    main = names(crr.model$coef)[j],
    xlab = "Failure time",
    ylab = "Schoenfeld residuals")
}
```



Residuos de Schoenfeld contra el tiempo de falla para cada covariable. Se observa que los residuos siguen una distribución no constante a lo largo de los tiempos de falla, lo que indica una posible violación del supuesto proporcional.

```
crr.time2<-crr(Melanoma$time,Melanoma$status,cov1=cov,cov2=cov[,3],tf=function(t) t)
crr.time2
```

```
## convergence: TRUE
## coefficients:
##      sexHombre      age invasionlevel.1 invasionlevel.2  cov[, 3]1*tf1
##      0.6291000    0.0045370    0.1627000    1.3530000    0.0007074
## standard errors:
## [1] 0.2726000 0.0090520 0.5064000 0.3874000 0.0003326
## two-sided p-values:
```

```
##          sexHombre          age invasionlevel.1 invasionlevel.2  cov[, 3]1*tf1
##          0.02100          0.62000          0.75000          0.00048          0.03300

crr.time3<-crr(Melanoma$time,Melanoma$status,cov1=cov,cov2=cbind(cov[,3], cov[,3]),tf=function(t) cbind
crr.time3

## convergence:  TRUE
## coefficients:
##          sexHombre          age
##          6.301e-01          4.561e-03
##          invasionlevel.1      invasionlevel.2
##          2.392e-01          1.354e+00
##  cbind(cov[, 3], cov[, 3])1*t cbind(cov[, 3], cov[, 3])2*tf2
##          5.703e-04          4.547e-08
## standard errors:
## [1] 2.726e-01 9.070e-03 7.545e-01 3.866e-01 1.127e-03 3.388e-07
## two-sided p-values:
##          sexHombre          age
##          0.02100          0.62000
##          invasionlevel.1      invasionlevel.2
##          0.75000          0.00046
##  cbind(cov[, 3], cov[, 3])1*t cbind(cov[, 3], cov[, 3])2*tf2
##          0.61000          0.89000
```

El modelo incluye un término de interacción con función de tiempo lineal, que muestra que el término de interacción (tiempo * nivel de invasión1) es estadísticamente significativo ($P = 0.033$).

Referencia

Zhang, Z. (2017). Survival analysis in the presence of competing risks. *Annals of translational medicine*, 5(3).